https://doi.org/10.5194/tc-2021-282
Preprint. Discussion started: 21 September 2021
© Author(s) 2021. CC BY 4.0 License.

The Cryosphere
Discussions
Open Access

Probabilistic Gridded Seasonal Sea Ice Presence Forecasting using Sequence to Sequence Learning

Nazanin Asadi[1], Philippe Lamontagne[2], Matthew King[2,3], Martin Richard[2], and K Andrea Scott[1]

[1]Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
[2]Ocean, Coastal and River Engineering Research Centre, National Research Council Canada, Ottawa, Canada
[3]Memorial University of Newfoundland, Newfoundland and Labrador, Canada.

Correspondence: Nazanin Asadi (n2asadi@uwaterloo.ca)

Abstract. Accurate and timely forecasts of sea ice conditions are crucial for safe shipping operations in the Canadian Arctic and other ice-infested waters. Given the recent observations on the declining trend of Arctic sea ice extent over the past decades due to global warming, machine learning (ML) approaches are deployed to provide accurate short-term to long-term forecasting. This study unlike previous ML approaches in the sea-ice forecasting domain, provides a daily spatial map of probability of ice in the study domain up to 90 day lead time. The predictions are further used to predict freeze-up/breakup dates and show their capability to capture these events within a valid time period (7 days) at specific locations of interest to communities.

1 Introduction

Sea ice presence is an important variable for northern communities and shipping operators. Sea ice forecasting needs to be carried out at various spatial and temporal scales to address different requirements of stakeholders. Short-term forecasts at high spatial resolution are important for day-to-day operations and weather forecasting (Carrieres et al., 2017), whereas longer term (eg. 60-90 day) forecasts are desired by shipping companies and offshore operators in the Arctic for strategic planning (Melia et al., 2016). In this study we are interested in these longer term forecasting methods. Typical approaches are usually statistical or dynamical in nature.

Seasonal sea ice forecasting using dynamical forecast models, that can propagate the state forward in time, is relatively new in comparison to weather forecasting, beginning with the study by Zhang et al. (2008) in which they evaluated the ability of an ensemble approach to predict the spring and summer Arctic sea ice extent and thickness for the summer of 2008. The majority of studies on sea ice prediction and forecasting focus on the pan-Arctic domain. A comparison between pan-Arctic and regional forecast skill was carried out by Bushuk et al. (2017), in which it was shown that the skill of seasonal forecasts in specific regions was dependent on the region and forecast month. This is due to regional differences of freeze-up and break-up dates, and the various environmental controlling factors.

Dynamical forecast models solve differential equations governing the physics of the underlying system. Solution methods for these types of equations are well-known and relatively robust. A key challenge with these models in operational forecasting is the high level of computational resources required to generate a forecast. This is a key advantage of statistical approaches such as multi-linear regression (Drobot et al., 2006) or canonical correlation analysis (Tivy et al., 2011). These approaches both determine a linear relationship between a set of predictor variables and a set of predictands.

More recently, convolutional neural networks (CNNs), which are able to learn nonlinear relationships between spatial patterns in input data and predictands, have been used to perform sea ice concentration prediction (Kim et al., 2020). The study by Kim et al. (2020) used eight predictors composed of a sea ice concentration data and variables from reanalyses to train 12 individual monthly models, and produced monthly spatial maps of sea ice concentration (SIC). Their results are compared against anomaly persistence and a random forest model. Their predictions were in good agreement with the mean September sea ice extent from 2017, with predictions from both the random forest and CNN model close to the observed ice extent. Similar to Kim et al. (2020), Horvath et al. (2020) focused on a monthly sea ice statistic, in this case the September sea ice minimum. This study used a Bayesian logistic framework to predict both a monthly average sea ice concentration and an uncertainty using atmospheric and oceanic predictor variables and sea ice concentration from satellite data. It was found the uncertainty was higher at the ice edge, although further analysis of this output was not given. In another study (Fritzner et al., 2020) two machine learning (ML) approaches, K-nearest-neighbour (KNN) and fully convolutional neural networks, have been compared with ensemble data assimilation at shorter time scales.

A recent approach that is closer to what is proposed here is IceNet (Andersson et al., 2021), which applies an ensemble of CNNs, each with a U-Nets architecture, to produce monthly maps of ice presence (probability SIC > 15%) for the next 6 months. Input to this model consists of climate variables, which is similar to other studies. A novel aspect is the training protocol, which consists of pre-training each ensemble member using a long time series of data from the Coupled Model Intercomparison Project phase 6 (CMIP6) and then fine-tuning the trained CNNs using SIC observations, followed by temperature scaling to produce a calibrated probability of ice.

None of the previously proposed ML approaches are able to produce a spatial forecast that propagates forward in time, in order to generate a spatiotemporal forecast. In this study we investigate a sequence-to-sequence learning approach to provide a spatiotemporal forecast of the probability of sea ice presence at daily time scale over the region of Hudson Bay, with forecast lead times up to 90 days. The method is similar to operational forecasting studies (Chevallier et al., 2013; Sigmond et al., 2013) except we are using a data-driven statistical approach, as compared to a physics-based model, and our forecasted variable is a number between 0 and 1 that indicates an (uncalibrated) probability of ice at a grid location, as compared to sea ice concentration.

2 Data

The present study utilizes ERA5 reanalysis for model predictors and validation. ERA5 is a recent reanalysis produced by the European Center for Medium Range Weather Forecasting. It consists of an atmospheric reanalysis of the global climate providing estimates of a large number of atmospheric, land and oceanic climate variables. The spatial resolution is ≈31 km and reanalysis fields are available every hour from 1979 - present (Wang et al., 2019). Observations are assimilated into the atmospheric model using a 4D-Variational data assimilation scheme. In this study ERA5 reanalysis data from 1985-2017 is used.

The sea ice concentration data used in ERA5 is from the EUMETSAT Ocean and Sea Ice Satellite Applications Facility (OSI-SAF) 401 dataset (Tonboe et al., 2016). These data are produced using a combination of passive microwave sea ice concentration retrieval algorithms to benefit from low sensitivity to atmospheric contamination of the surface signal, while maintaining an ability to adapt to changes in surface conditions through the use of variable tie-points for ice and water (Tonboe et al., 2016). Although the SIC is gridded to a 10 km grid, the spatial resolution of these data is limited by the instrument field of view of the 19 GHz channel used in the SIC retrieval, which is 45 km × 69 km. When the SIC data are ingested into ERA5 the SIC values that are less that 15% are set to zero, and SIC is set to zero if SST is above a specified threshold.

The current study utilizes the following input variables from ERA5 dataset over the period of 1985-2017: sea ice concentration, sea surface temperature, 2m temperature (t2m), surface sensible heat flux, wind 10 meter U-component (u10), wind 10 meter V-Component (v10) and landmask and additive degree days (ADD) derived from the t2m variable. All the input variables except sea ice concentration and landmask are normalized before being input to the network.

3 Study Region

For the present study we focus on the Hudson Bay System, consisting of Hudson Bay, James Bay, Hudson Strait and Foxe Basin. The area is bordered by 39 communities, where 29 of them are exclusively accessible by sea or air. These communities rely extensively on sealift operations during the ice-free season to receive their yearly resupply of fuel and goods too heavy to be flown. Shipping traffic, mostly confined during the ice-free and shoulder season, is also generated by mining, fishing, tourism and research activities (Andrews et al., 2018).

The study area is seasonally covered by first-year ice, with open water over most of the domain each summer, with the exception of some small regions in Foxe Basin (the northern portion of the study region). The seasonal cycle of ice cover in this region is dominated by local atmospheric and oceanic drivers (Hochheim and Barber, 2014). Freeze-up generally starts in November (earlier in the northern part of the region) and lasts for a couple of months. Break-up usually starts in May or June, and the break-up period is a little longer than freeze-up, at 2-3 months. Recent decades show earlier breakup and later freeze-up. The trends and their significance are dependent on the region (Hochheim and Barber, 2014; Andrews et al., 2018).

4 Forecast model architecture

The medium-term forecasting problem of this study can be formulated as a spatiotemporal sequence forecasting problem that can be solved under the general sequence-to-sequence (Seq2Seq) learning framework in machine learning domain (Sutskever et al., 2014). In Seq2Seq learning, which have successful applications in machine translation (Cho et al., 2014), video captioning (Venugopalan et al., 2015), speech recognition (Chiu et al., 2018), etc, the target is to map a sequence of inputs to a sequence of outputs, where the inputs and outputs can be different lengths. The architecture of these models normally consist of 2 major components; encoder and decoder. The encoder component transforms a given input to an encoded state of fixed shape, while the decoder part takes that encoded state and generates an output sequence with the desired length.

For this study, following the encoder-decoder architecture explained above, two spatiotemporal sequence models are developed where the second one is the Augmented version of the Basic one. For both models, the prediction sequence is unrolled over a user-specified number of forecast days to produce ice presence probability forecasts on a spatial grid each day with a scale of $\sim$ 31 km (the same as input).

4.1 Basic model

The encoder section of the Basic model takes as input the last three days of environmental conditions. Each input sample is of size $(3 \times W \times H \times C)$ where 3 is the number of historical days, W and H are the width and height of the raster samples in their original resolution and C is the total number of input variables (here 8).

The encoder starts with passing each daily sample through a feature pyramid network (Lin et al., 2017) so as to detect environmental patterns at both the local and large scales. Next, the sequence of extracted feature grids are further processed through a convolutional LSTM layer (ConvLSTM) (Hochreiter and Schmidhuber, 1997; Xingjian et al., 2015), returning the last output state. This layer learns a single grid representation of the time series that also preserves spatial locality. Finally, the most recent day of historic input data is concatenated with the ConvLSTM output. The encoder provides as output a single raster with the same height and width but higher number of channels such as to represent the fully encoded system state.

The final encoded state is then fed to a custom recurrent neural network (RNN) decoder that extrapolates the state across the specified number of time-steps. It takes as input the encoded state with multiple channels and as output produces a state with the same width and height and desired time-steps.

The custom RNN decoder, as is common of many RNN layers, maintains both a cell state and a hidden state (Yu et al., 2019). First, the initial cell state and hidden state are initialized with the input encoded state. Then, at each time-step and for each of the states, the network predicts the difference, or residual, from the previous state to generate the updated states using depthwise separable convolutions (Howard et al., 2017). The output of the decoder section is the concatenation of the cell states from each time-step.

Finally, a time-distributed network-in-network (Lin et al., 2013) structure is employed to apply a 1×1 convolution on each time-step prediction to keep the grid size the same but reduce the number of channels to one, representing the daily probabilities of ice presence over the forecast period (e.g. up to 90 days).

4.2 Forecast-Augmented model

A slight variant of the Basic model is developed so as to accept a second input. This second input has the same height and width as the first input but corresponds to climate normal of three variables over the required period (e.g., 60 or 90 days), where these variables are t2m, u10 and v10 and their climate normal is calculated from 1985 to the last training year for each forecast day. The original encoder structure for historical input data remains unchanged, but the primary feature pyramid network in the Basic model is joined by a secondary feature pyramid network from the second input. A secondary variant of the decoder component is implemented which accepts this encoded forecast sequence in order to produce superior estimates of the residuals at each of the future time-steps. Here, the decoder is designed in a way that the number of forecast input time-steps can be either equal or less than the forecast length. This network is referred to as the Augmented model.

5 Description of Experiments

Since the overarching goal is to provide a tool to stakeholders that can be used operationally, it required a training and validation protocol that truly assess the forecasting skills without using future data. After extensive experimentation the following protocol was found to lead to the best results.

For each month of a year a separate model is trained on data from the given month as well as the preceding and following month. For example, the 'April model' is trained using data from March 1 to May 31. This monthly model is initially trained on data from a fixed number of years, chosen to be 10 years. After this initial experiment, to predict each following test year i, using a rolling forecast prediction, the model from year $i-1$ is retrained with data from year $i-2$ and also, data from year $i-1$ is used as validation for early stopping criteria and to evaluate the training performance. For example, if the initial model is trained on 10 years, data from year 11 is used as validation and first predictions are launched at year 12. The model for year 12 is then retrained with data from year 11 and validated on year 12 to predict year 13 and so on. Thereby, the output's statistics are calculated on forecasts of 1996 to 2017. Since the retraining process only uses data of one year for training and validation, it is computationally fast and efficient.

The ML models are implemented using the TensorFlow Keras open-source library with stochastic gradient descent (SD) optimizer with learning rate of 0.01, momentum of 0.9 and binary cross-entropy loss function. The maximum training epoch for the initial model and the retraining process is 60 and 40 respectively and for both cases the training process stops if the validation accuracy is not improved after 5 epochs.

In order to evaluate the performance of the neural network model, the results are compared with climate normal. This is defined as the average of ERA-5 sea ice concentration (thresholded at 15%) from 1985 to the last year in the training set for each experiment. While inputs of each model in training and test procedure are coming from 3 months of year, only the results from the central month (2nd of 3) is selected for evaluating the results of that model in the following section.

6 Results

6.1 Presence of Ice Forecasts

Given that our models predict spatial maps of sea ice presence probability over a grid at a spatial resolution of 31 km, we first apply a 50% threshold to this probability to convert each pixel to ice or water. From this, for each day in the test set, we have 90 binary accuracy maps of Hudson Bay. To summarize these results, in Fig 1(a,b,c) we show the model binary accuracy as a function of forecast lead day for each month. For example, the top row of Fig 1b shows the accuracy of forecasts launched in January using Basic model for forecast lead days of 1 to 90. E.g., the first top-left box in this figure (Fig 1(b)) corresponds to the average accuracy after 1 day forecast for all forecasts launched between January 1 and January 31, ending in January 2 to April 1 and the second box corresponds to average accuracy of forecasts launched between January 1 and January 31 ending in January 3 to April 2.

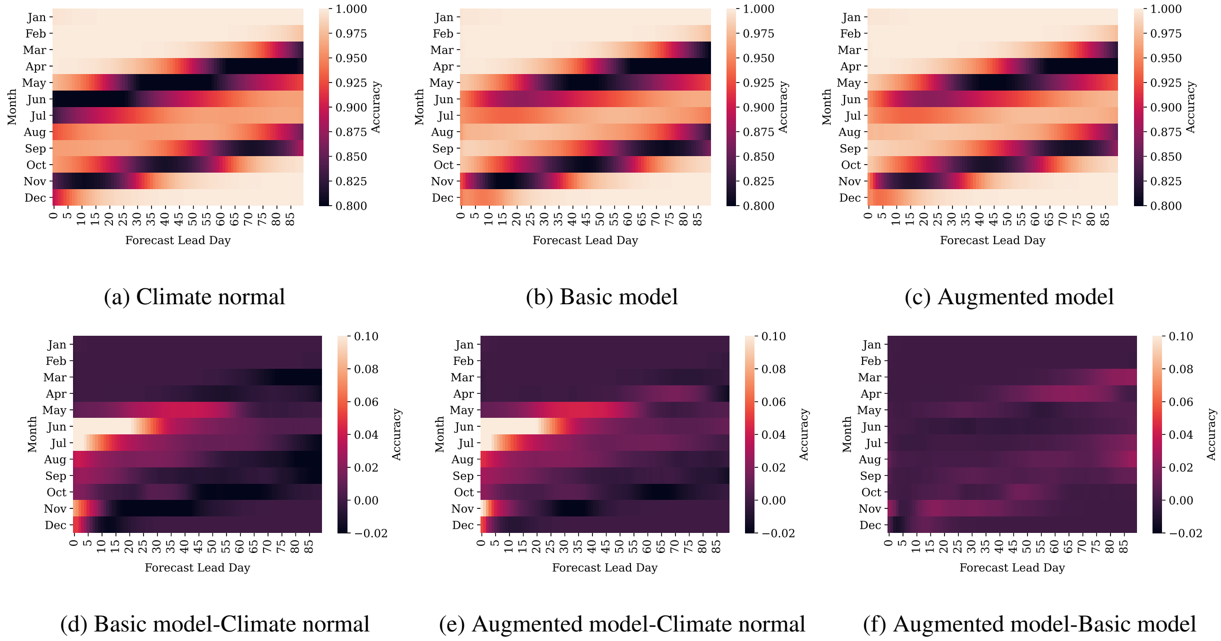
The accuracy is very close to 100% for the month on January and for these lead times which cover the months of January, February and March, as would be expected, because at this time the region is consistently covered with ice. In contrast, for forecasts of June and July, the beginning of the open water season, the climate normal struggles to accurately capture the ice cover for lead times of 1 to 50 days likely due to inter-annual variability and the impact of climate changes. However, the Basic and Augmented models proposed here have significantly higher accuracies than climate normal over these months (Fig 1d), especially in early lead time, with no degradation at longer lead times.

We also note improvements in proposed models at early lead times for August, September, October and November, as compared to climate normal. If we recall freeze up starts in mid-October or November in the study region and lasts for approximately two months (Hochheim and Barber, 2014), we note these forecasts correspond to the freeze-up period. When comparing the Augmented model with Climate normal (Fig 1e) and with the Basic model (Fig 1f), improvements in accuracy can be seen in particular for longer lead times in March/April and July/August, and improvements at shorter lead times (15-50 days) for November. Using additional climate variables for the input of the Augmented model is showing its impact here where in the periods that Basic model is worse than climate normal (Fig 1d), the Augmented model has better accuracy and is closer accuracy to climate normal. For example, March forecasts at 80 to 90 lead day (Fig 1e, Fig 1f). However, in general from Fig 1d and Fig 1e it can be seen that both Basic and Augmented models do not lead to degraded forecasts in comparison to climate normal for any months or lead times (the largest reduction in accuracy is 0.02). To evaluate the probabilistic accuracy of the Basic and Augmented models, the Brier score (Ferro, 2007) is calculated for each predicted probability map using the following formula,

$$BS = \frac{1}{N}\sum_{t=1}^{N}(P_t - O_t)^2, \tag{1}$$

where t and N represent the index of the grid cell and total number of grid cells in the probability map, respectively. Also, P_t is the model prediction and O_t is the corresponding observation at grid cell t. Figure 2 represents the monthly averaged Brier scores for Basic model (Fig 2a), Augmented model (Fig 2b) and their differences (Fig 2c) as a function of lead day. Similar to Fig 1, each value at index (i,j) of panel (a) and (b) of Fig 2 represents the average Brier score of all predictions in the test set that are launched at month i at lead day j where $1 \leq i \leq 12$ and $1 \leq j \leq 90$. The pattern observed is similar to that for binary accuracy as the Brier score is higher in freeze-up and breakup periods for both models and their difference (2c) indicates a better score for the Augmented model at longer lead days especially for March, Apr, July and August. In contrast for some cases like 60-90 lead day forecasts of September model the Brier score of the Basic model is around 0.01 better than the Augmented model.

The calibration curves of the Basic and Augmented September models are shown in Figure 3. These curves represent the observed frequency of ice presence, where the frequency is calculated over the entire domain, versus the forecasted probabilities for different lead days of forecasts launched in September. For early lead days both models, especially the Basic model, show



(a) Climate normal

(b) Basic model

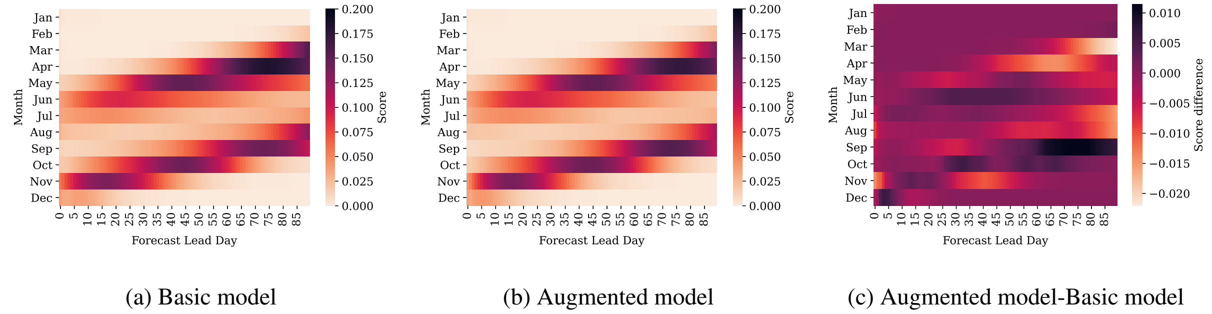
(c) Augmented model

(d) Basic model-Climate normal

(e) Augmented model-Climate normal

(f) Augmented model-Basic model

Figure 1. Model performance and improvements as a function of lead time. Top row panels describe performance of each model (a)-(c) while bottom row represents the accuracy differences between the models (d-f). The Augmented model is trained with additional 90-day climate normal input. Most differences are observed in breakup and freeze-up seasons.



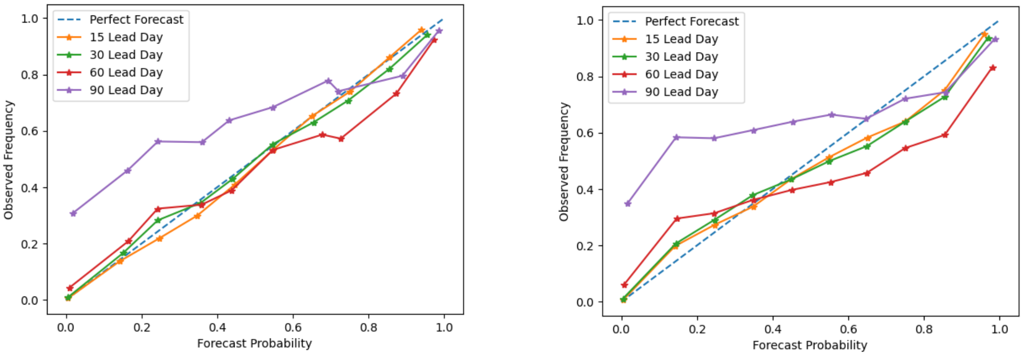
(a) Basic model

(b) Augmented model

(c) Augmented model-Basic model

Figure 2. Brier score of the Basic and Augmented model as a function of lead time and their score difference. Most differences are observed in breakup and freeze-up seasons.

185 close to perfect calibration (blue line) but at 60 lead day the underestimation is more significant for the Augmented model with lower forecasted probabilities of ice in comparison to observations, while at 90 lead day the overestimation is more significant for the Augmented model, with a forecasted probability that is much higher than the observed frequency of ice. This suggests that in comparison to observations, freeze up may be delayed at 60 lead days for regions with freeze-up dates around November and may be too early at 90 lead days for regions with freeze-up date around December for the Augmented model.



(a) Basic model

(b) Augmented model

Figure 3. Calibration curve of the Basic and Augmented model for forecasts initiated from September 1st to 30th for different lead days. Both models are underestimating the probability at 90 lead day when observed frequencies are less than 0.8 and overrestimating when a probability of greater than 0.8 is expected.

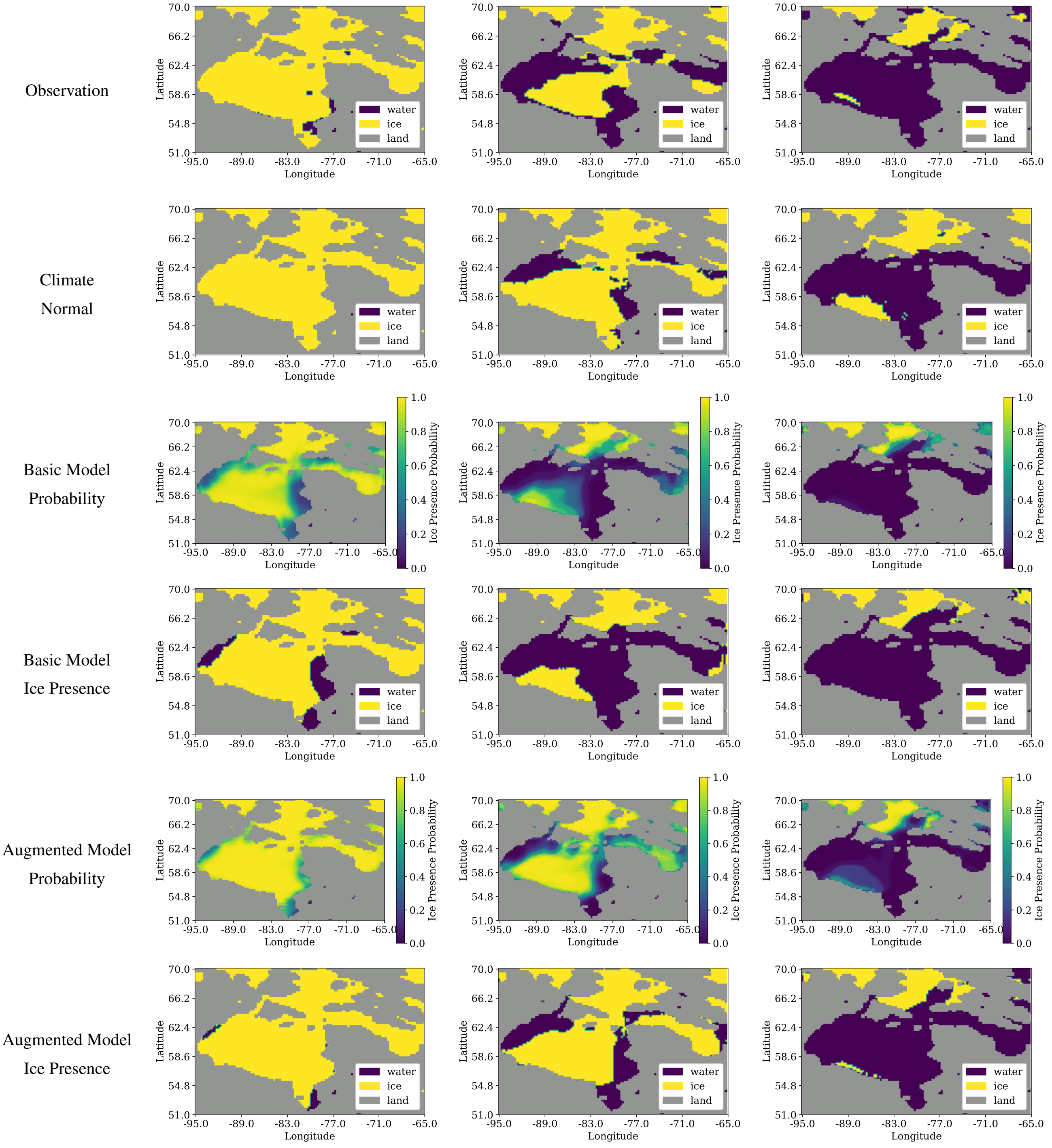
Monthly averaged accuracies do not provide information about model performance at each location in the spatial domain, or at a finer time scale. The model proposed here, unlike many previous ML approaches in the sea-ice forecasting domain, provides a spatial map of the probability of ice for each day in the forecast period. In Fig 4 the spatial distribution of ice and water is shown with the probability of ice for three different dates during the breakup period. The observations (top row) are ice and water obtained by applying a threshold of 15% to SIC from ERA5 for the given data. For example, given that the forecasts are launched on May 6th 2014, the left column (after 30 days) corresponds to the sea ice state on June 5th 2014.

The forecast after 30 days indicates both the Basic and Augmented models predict a reduction in ice cover along the east coast of Hudson Bay which is in better agreement with observations than climate normal. Similarly, after 50 and 70 days, climate normal has more ice and and Basic model has less ice in the central part of Hudson Bay, while Augmented model is in better agreement with observations. In comparing the probability maps for Basic and Augmented model, it can be noted that the Basic model has reduced ice presence probability in the southern part of the domain and increased ice presence probability in the northern part of the domain, as compared with Augmented model. Overall we find the agreement in spatial pattern of break up to be in good agreement with the Observations, in particular for the Augmented model.

6.2 Assessment of operational capability

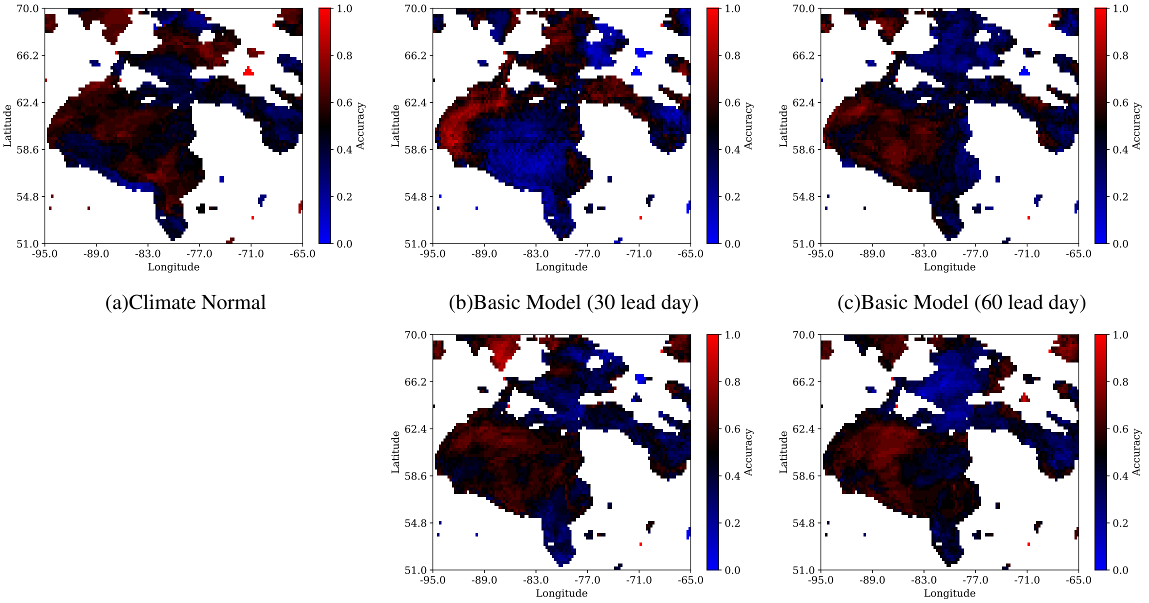
The accuracy of the model in predicting freeze-up and breakup date is indicative of operational capability of the trained models to support shipping operations during the shoulder season. Following the definition used by the Canadian Ice Service (CIS), the freeze-up date of each pixel in a year is the first date in the freeze-up season (Oct 1st to Jan 31st for Hudson Bay) that ice (value of 1 after thresholding the predicted probabilities at 50%) is observed for 15 continuous days. A similar procedure is carried out to predict breakup, with the exception that the pixel must be considered water (value of 0 after thresholding its predicted probability) for 15 continuous days in a row for breakup to have occurred in the breakup season (May 1st to July 31st

https://doi.org/10.5194/tc-2021-282
Preprint. Discussion started: 21 September 2021
© Author(s) 2021. CC BY 4.0 License.



(a) After 30 days (b) After 50 days (c) After 70 days

Figure 4. An illustration of the May models capturing the spatial patterns of Breakup. The models are trained starting on May, June, and July. The forecasts are started at May, 6th, 2014 and are displayed after 30 days (June 5th), 50 days (June 25th) and 70 days (July 15th).



(a)Climate Normal

(b)Basic Model (30 lead day)

(c)Basic Model (60 lead day)

(d)Augmented Model (30 lead day)

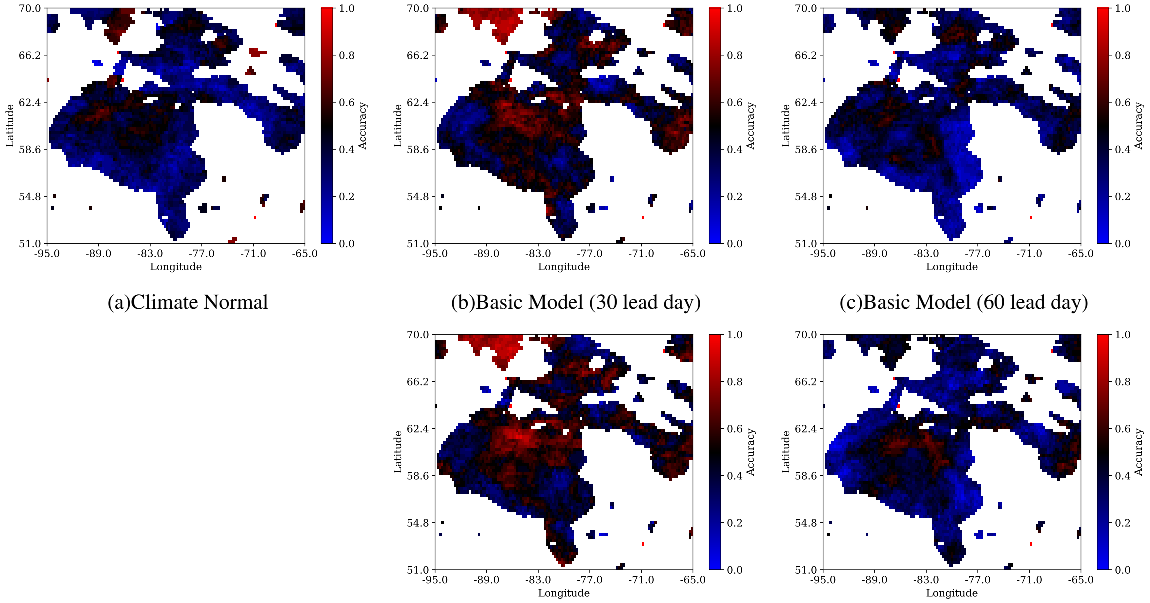
(e)Augmented Model (60 lead day)

Figure 5. Accuracy of predicted freeze-up date within 7 days.

for Hudson Bay). These per pixel per year freeze-up/breakup dates are calculated for observations, climate normal and model predictions at 30 and 60 lead day. A similar definition of freeze up and break up is used in Sigmond et al. (2016).

Figure 5 and 6 are showing the overall accuracy map of the climate normal as well as the Basic and Augmented model for 30 and 60 lead days. To obtain each accuracy map, first the predicted and observed freeze-up/breakup dates per pixel per year are compared. If the 2 dates are within 7 days of each other, the prediction is correct. Then, the results are averaged over the total number of years to get a value between 0 and 1. The freeze-up accuracy maps at Fig 5 show that except the Basic model's prediction at 30 lead day (Fig 6b), other maps are showing similar patterns of accuracy. The difference between Basic model's freeze-up prediction map at 30 lead day and other maps is due to its difference in accuracy of ice presence forecasts. The freeze-up event in the central region of Hudson Bay is observed in December. For that purpose, we should look at the Basic and Augmented model's forecasts from November and October model for 30 and 60 lead day respectively. It was found (not shown) that the December ice presence accuracy of the Basic model at 30 lead day have lower accuracy in central region and higher accuracy in Hudson Strait comparing to other methods, which explains the difference in freeze-up prediction maps.

The breakup prediction ability of the models are presented in Fig 6. In contrast to freeze-up, the climate normal (fig 6a) is showing an overall poor accuracy in prediction of breakup dates over the spatial domain. The Augmented model at 30 lead day (Fig 6d) has the best accuracy, especially in the central regions, while the breakup prediction accuracy degrades at 60 lead day for both the Basic and Augmented models. The Augmented model's breakup prediction accuracy at 60 lead day (Fig 6e) is slightly higher than that of the Basic model at 60 lead day (Fig 6c) in the central zone.



(a)Climate Normal

(b)Basic Model (30 lead day)

(c)Basic Model (60 lead day)

(d)Augmented Model (30 lead day)

(e)Augmented Model (60 lead day)

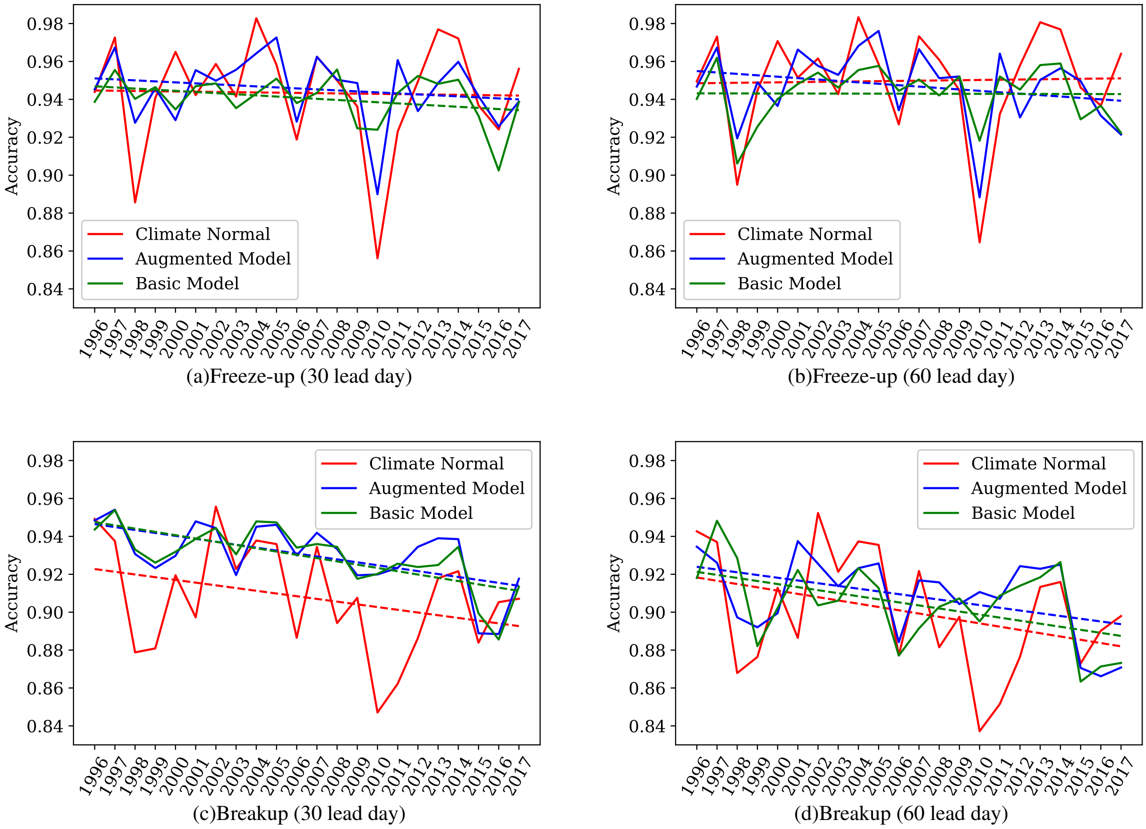
Figure 6. Accuracy of predicted breakup date within 7 days

The variability of the models' accuracy in freeze-up (Oct1st to Jan 31st) and breakup (May 1st to July 31st) predictions is represented in Fig 7 for 30 and 60 lead day. For each prediction, its trend is also shown by a line with the same color. While no significant trend is observed for freeze-up season accuracy at both lead days, the breakup plots ((c) and (d)) show a declining trend of 2% in models and climate normal accuracy over the years. Also, similar to Fig 6 for freeze-up/breakup date predictions, Augmented and Basic models have their highest improvement comparing to climate normal for breakup season at 30 lead day. In addition, for both cases, 2010 is showing an extreme case where climate normal has the lowest accuracy over the entire period. For that year, the Augmented model is also showing lower freeze-up season accuracy comparing to other years but its accuracy over breakup season doesn't show any significant variability over the years.

The ability of the model to predict freeze-up and breakup dates can provide helpful information for local communities and shipping operators. Here, the nearest pixels to three sample ports, Churchill, Inukjuak, and Quataq, are selected and the freeze-up/breakup date predictions of the models at 30 and 60 lead day versus the observed dates are presented in figures 8 and 9. The red line in each plot represents a perfect one-to-one prediction and the pink region is showing the acceptable 7 days difference that will still be considered as a correct prediction according to the CIS criteria. The width of the pink zone on each plot varies as the total time frame of breakup and freeze-up at each location is different. In addition, the year of 2010 is omitted from these plots as it was an anomalously warm year (Hochheim and Barber, 2014).

For freeze-up, in Fig 8, 30 lead day predictions are more concentrated and closer to the pink zone while there is more dispersion and outliers observed for 60 lead day predictions. In addition, Augmented model predictions have fewer outliers than Basic model predictions. For the port of Churchill, predictions are close to the center and inside or close to the pink zone

https://doi.org/10.5194/tc-2021-282
Preprint. Discussion started: 21 September 2021
© Author(s) 2021. CC BY 4.0 License.



(a)Freeze-up (30 lead day)

(b)Freeze-up (60 lead day)

(c)Breakup (30 lead day)

(d)Breakup (60 lead day)

Figure 7. Accuracy of the climate normal, Basic model and Augmented model freeze-up and breakup predictions over the years at 30 and 60 lead day. Dotted lines show the trend.

for both models both lead days compared to other locations. The Basic model especially at 30 lead day, predicts freeze-up dates of several years with a consistent delay for Inukjuak while for Quataq its predictions are earlier than observed dates. In Fig 9, similar to freeze-up, breakup dates are better captured by Augmented model at 30 lead day as compared to 60 lead day, where predictions are more scattered. Also, the patterns of early and delayed predictions are not as visible as freeze-up for breakup dates for Inukjuak and Quataq ports.

7 Discussions and Conclusion

This study has focused on sea ice presence probability forecasting using deep learning methods at a daily time scale with lead times up to 90 days in advance. The Basic model uses 8 input variables from the ERA5 dataset over the last 3 days before the initial date on which the forecast is launched. To improve the Basic model performance, the Augmented version is proposed where it takes an additional input from climate variables over the forecasting period. Compared to operational forecasting

© Author(s) 2021. CC BY 4.0 License.

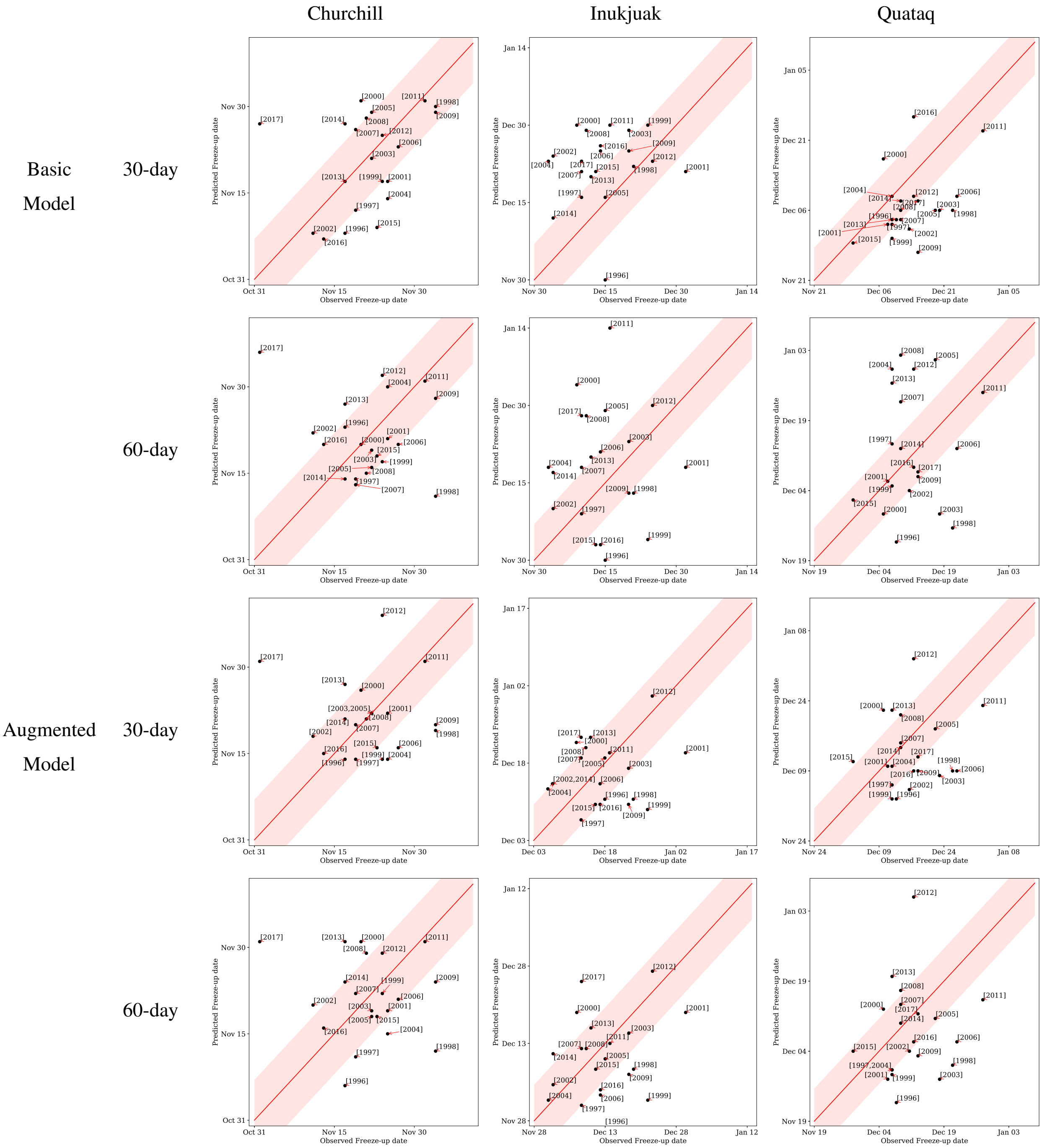


Figure 8. Freeze-up dates of sample ports at 30 lead day and 60 lead day forecast versus the observations. The red line represents perfect predictions and the pink area represents +/- 7 days of the red line, which is commonly assumed as an acceptable error range.

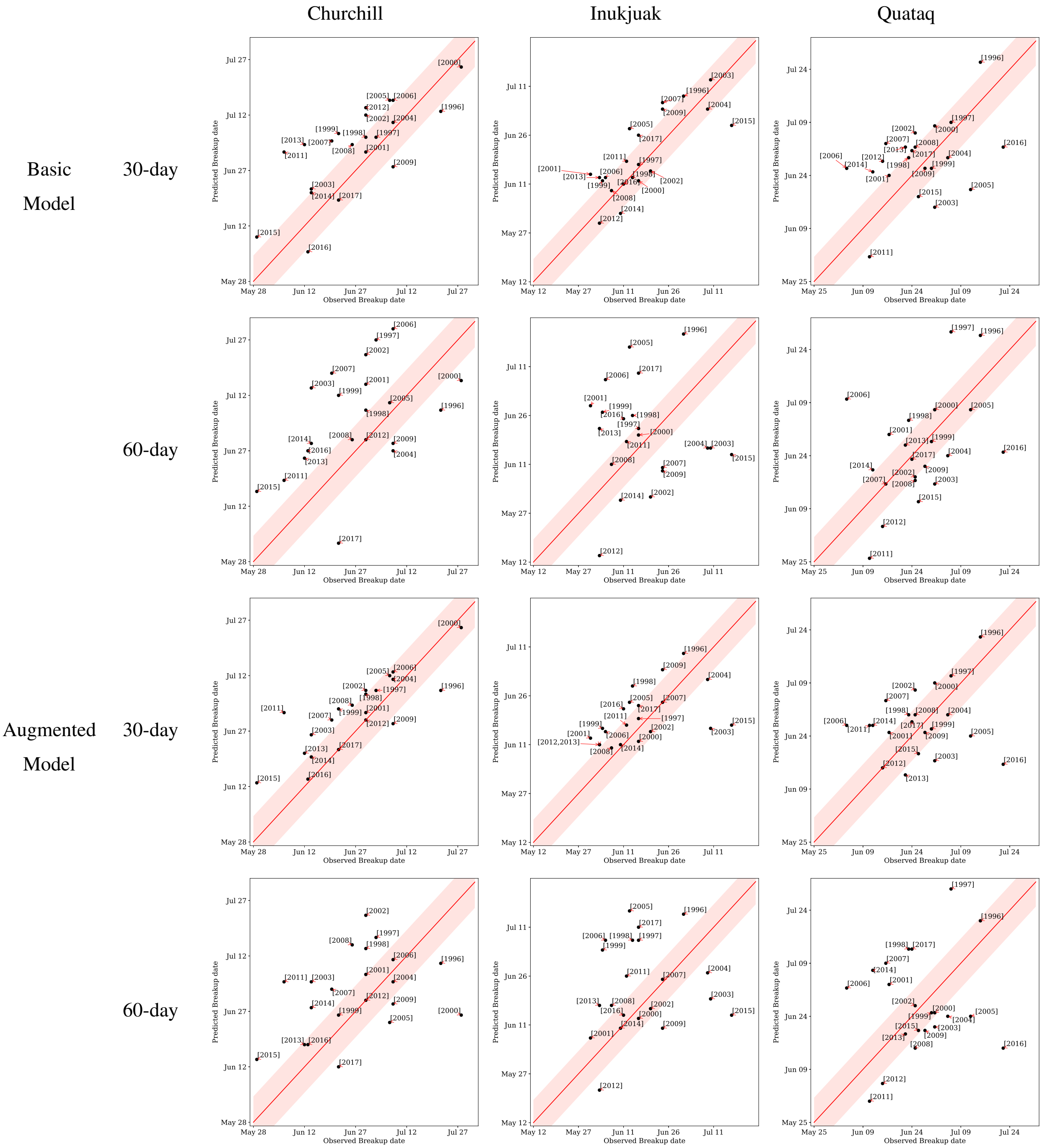


Figure 9. Breakup dates of sample ports at 30-day and 60-day forecast versus the observations. The red line represents perfect predictions and the pink area represents +/- 7 days of the red line, which is commonly assumed as an acceptable error range.

systems in this domain, the proposed approach has the advantage of time efficiency as once the initial model is trained, the fine-tuning process for new inputs (consisting of one year of training data) takes around 15 minutes on a Tesla GPU and each inference takes around 10 seconds to complete.

Comparing the binary accuracy of the two proposed models and climate normal demonstrated improvements of up to 10% relative to climate normal for both the breakup and freeze-up seasons, especially for early lead days (up to 30 days). The probability assessment by the calibration analysis and Brier score also revealed most differences in breakup and freeze-up season with the Augmented model showing better scores comparing to Basic model.

The analysis on breakup and freeze-up date prediction of the models shows the Augmented model is more capable at accurately predicting these dates within 7 days compared to the Basic model while the accuracy of both models degrades when increasing the lead day. It should be noted that both models have substantial improvement over climate normal at 30 lead day for breakup date prediction.

Breakup and freeze-up date prediction analysis at sample ports shows that freeze-up date predictions of both models are less disperse and are close to accepted +/-7 days zone while for breakup date prediction the observations and predictions show more diversity over the analyzed test years.

As future work, we plan to expand the experiments over the entire Arctic region, and deploy ensemble methods using more recent deep learning architectures. Looking into possible improvements by adding SIC anomaly as additional input variable as investigated by Kim et al. (2020) is another path to explore.

Author contributions. PL, MK and MR designed and initiated the study and proposed the model. NA designed the experimental setup and performed the simulations and analysis of the results. PL and KAS supervised the study and provided feedback. NA, PL, MK, and KAS contributed to the development and writing of this paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors would like to acknowledge funding from the National Research Council of Canada through the AI4Logistics and Ocean Programs and computing resources provided by Compute Canada.

References

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., et al.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature Communications, 12, 2021.

Andrews, J., Babb, D., Barber, D. G., and Ackley, S. F.: Climate change and sea ice: shipping in Hudson Bay, Hudson Strait, and Foxe Basin (1980–2016), Elementa: Science of the Anthropocene, 6, 2018.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Gudget, R., Rosati, A., and Yang, X.: Skillful regional prediction of Arctic sea ice on seasonal time scales, Geophysical Research Letters, 44, 10.1022/2017/GL073 155, 2017.

Carrieres, T., Buehner, M., Lemieux, J.-F., and Pedersen, L., eds.: Sea ice analysis and forecasting : towards an increased resilience on automated prediction system, Cambridge University Press, 2017.

Chevallier, M., Salas y Mélia, D., Voldoire, A., and Deque, M.: Seasonal forecasts of pan-Arctic sea ice extent using a GCM-based seasonal prediction system, Journal of Climate, 26, 6092–6104, 2013.

Chiu, C.-C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R. J., Rao, K., Gonina, E., et al.: State-of-the-art speech recognition with sequence-to-sequence models, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4774–4778, IEEE, 2018.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078, 2014.

Drobot, S., Maslanik, J., and Fowler, C.: A long-range forecast of Arctic summer sea-ice minimum extent, Geophysical Research Letters, 33, L10 501, 2006.

Ferro, C. A.: Comparing probabilistic forecasting systems with the Brier score, Weather and Forecasting, 22, 1076–1088, 2007.

Fritzner, S., R., G., and Christensen, K.: Assessment of high resolution dynamical and machine learning models for prediction of sea ice concentration in a regional application, Journal of Geophysical Research, Oceans, 2020.

Hochheim, K. P. and Barber, D.: An update on the ice climatology of the Hudson Bay System, Arctic, Antarctic and Alpine Research, 46, 66–83, 2014.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Horvath, S., Stroeve, J., Rajagopalan, B., and Kleiber, W.: A Bayesian logistic regression for probabilistic forecasts of the minimum September Arctic sea ice cover, Earth and Space Science, 7, e2020EA001 176, 2020.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861, 2017.

Kim, Y., Kim, H.-C., Han, D., Lee, S., and Im, J.: Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks, The Cryosphere, 14, 1083–1104, 2020.

Lin, M., Chen, Q., and Yan, S.: Network in network, arXiv preprint arXiv:1312.4400, 2013.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S.: Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125, 2017.

Melia, N., Haines, K., and Hawkins, E.: Sea ice decline and 21st century trans-Arctic shipping routes, Journal of Geophysical Research, 43, 9720–9728, 2016.

Sigmond, M., Fyfe, J., Flato, G., Kharin, V., and Merryfield, W.: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system, Geophysical Research Letters, 40, 529–534, 2013.

Sigmond, M., Reader, M., Flato, G., Merryfield, W., and Tivy, A.: Skillful seasonal forecasts of Arctic sea ice retreat and advance dates in a dynamical forecast system, Geophysical Research Letters, 43, 12–457, 2016.

Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in: Advances in neural information processing systems, pp. 3104–3112, 2014.

Tivy, A., Howell, S., Alt, B., Yackel, J., and Carrieres, T.: Origins and levels of seasonal skill for sea ice in Hudson Bay using canonical correlation analysis, Journal of Climate, 24, 1378–1394, 2011.

Tonboe, R., Eastwood, S., Lavergne, T., Sørensen, A., Rathmann, N., Dybkjaer, G., Pedersen, L., Høyer, J., and Kern, S.: The EUMETSAT sea ice concentration climate data record, The Cryosphere, 10, 2016.

Venugopalan, S., Rohrbach, M., Donahue, J., Mooney, R., Darrell, T., and Saenko, K.: Sequence to sequence-video to text, in: Proceedings of the IEEE international conference on computer vision, pp. 4534–4542, 2015.

Wang, C., Graham, R., Wang, K., Gerland, S., and Granskog, M.: Comparison of ERA5 and ERA-Interim near surface air temperature, snowfall and precipitation over Arctic sea ice: effect on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, 2019.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, pp. 802–810, 2015.

Yu, Y., Si, X., Hu, C., and Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures, Neural computation, 31, 1235–1270, 2019.

Zhang, J., Steele, M., Lindsay, R., Schweiger, A., and Morison, J.: Ensemble 1-year predictions of Arctic sea ice for the spring and summer of 2008, Geophysical Research Letters, 35, 1–5, 2008.